Research Article Impact Factor: 0.621 ISSN: 2319-507X
Neelam Gilorkar, IJPRET, 2014; Volume 2 (9): 628-637 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

IDENTIFICATION OF REAL TIME HAND GESTURE USING SCALE INVARIANT FEATURE TRANSFORM

NEELAM K. GILORKAR

Department of Electronics and Telecommunication, Government College of Engineering Amravati, India.

Accepted Date: 27/02/2014 ; Published Date: 01/05/2014

Abstract: This paper presents a vision based system that provides a feasible solution to Indian sign Language(ISL) and American Sign Language (ASL) identification of hand gestures of alphabet. The proposed system is advance and is meant to substitute interpreters, gloves or any other marker devices. Sign language recognition system typically elaborate three steps preprocessing, feature extraction and classification. The system uses digital image processing techniques using Scale Invariant Feature Transform (SIFT) algorithm and same is used to extract the features.

Keywords: American Sign Language (ASL) Feature extraction, hand gesture, Indian Sign Language (ISL), Scale Invariant Feature Transform (SIFT).

Corresponding Author: MS. NEELAM K. GILORKAR

Access Online On:

www.ijpret.com

How to Cite This Article:

PAPER-QR CODE

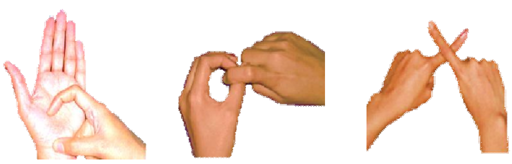
Neelam Gilorkar, IJPRET, 2014; Volume 2 (9): 628-637

-

INTRODUCTION

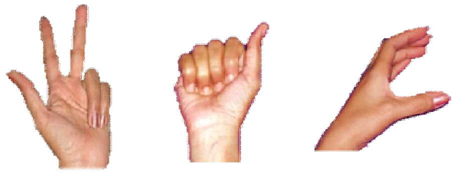
Sign language is widely used by people who cannot speak and hear or people who can hear but cannot speak. A sign language is composed of various gestures formed by different hand shapes, movements and orientations of hands or body, or facial expressions. Around the world, there are various sign languages each with its own vocabulary and regional dialects. These include American Sign Language (ASL) in Northern America, British Sign Language(BSL) in Great Britain, Japanese Sign Language(JSL) in Japan, South African Sign Language(SASL), etc. Gestures are used by the deaf people to express their thoughts. But the use of these gestures are always limited in the deaf-dumb community, normal people never try to learn the sign language. This causes a big gap in communication between the deaf–dumb people and the normal people. Usually deaf people seek the help of sign language interpreters for translating their thoughts to normal people and vice versa. But this system is very costly and does not work throughout the life period of a deaf person. So a system that automatically recognizes the sign language gestures is necessary. Such a system can minimize the gap between deaf people and normal people in the society.

By deaf community ISL is expressed by both hand gestures and ASL is expressed by single hand gesture as illustrated in figure 1. Finger spelling is used to form words with letter by letter coding. Letter by letter signing can be used to express words for which no sign exists, the words for which the signer does not know the gestures or to emphasis a particular word. So the recognition of the finger spelling has a key importance in sign language recognition. This paper present a method of automatic recognition system of static gestures in Indian and American sign language alphabet and numbers. The sign considered for recognition are 26 letters alphabets and the 0-9 numbers. Some ASL and ISL gestures are shown in Fig. 1.



(a) (b) (c)

-



(d) (e) (f)

Fig.1 ISL signs for (a) R, (b) Q, (c) X,

ASL signs for (d) 3, (e) A, (f) C

LITERATURE SURVEY

Recently, there has been a surge of interest on hand detection, tracking, and gesture recognition. While a comprehensive literature review of this field would be a daunting challenge because of the huge number of publications. Human hands are the parts that people use the most and with the most ease to interact with the world. Hands are highly articulated structures with some 27 degrees of freedom (DOF). Because of these high DOF of the human hand, hand gesture recognition is indeed an extremely challenging task.

Mainly, these systems are divided into two approaches, namely the glove-based and the vision-based approach[1]. The first category requires user to wear a sensor glove or a colored glove. The wearing of the glove simplifies the task of segmentation during processing. The Glove based approaches utilize sensor devices for digitizing hand and finger movements into multi-parametric data. The additional sensors facilitate detection of hand configuration and movement. However, the sensor devices are quite costly and wearing gloves or trackers is sore and enlarges the "time-to interface," or setup time. However, the data glove and its attached wires are still inconvenient and awkward for users to wear. Moreover, the cost of the data glove is often too expensive for regular users. The shortcoming of this approach is that the user has to wear the sensor hardware along with the glove during the operation of the system. On the other hand, vision-based methods are more natural and useful for real-time applications. Vision based approach uses image processing algorithms to detect and track hand signs as well as facial expressions of the user. This approach is easier to the user since there is no need to wear any extra hardware. However, there are accuracy problems related to image processing algorithms and these problems are yet to be modified.

Also, vision based systems need application-specific image processing algorithms, programming, and machine learning. From the perspective of the features used to represent

-

the hand, vision based hand tracking and gesture recognition algorithms are classified into two categories:

- 3D hand model-based approach
- Appearance based approach

3D hand model-based approach

This approach is very computational intensive and also, systems for live analysis are still to be advanced. However, 3D hand model method required a huge database with the entire characteristic shapes under several views due to deformable objects with many DOFs [4].

Appearance based approach

In this approach image features are extract to model the visual appearance of the hand and compare these features with the extracted features from the video frames as our approach. They have real time performance because of the easier 2-D image features that are used[2].

SYSTEM FUNCTIONALITY

The system is designed to visually recognize static signs of the Indian Sign Language (ISL) and American sign Language (ASL) all signs of ISL alphabets using bare hands. The user or signers are not required to wear any gloves or to use any devices to interact with the system. But, since different signers vary their hand shape size, and operation habit and so on, which bring more difficulties in identification. The entire method consists of the following three main stages: Image Preprocessing, Feature extraction using SIFT and Gesture Identification.

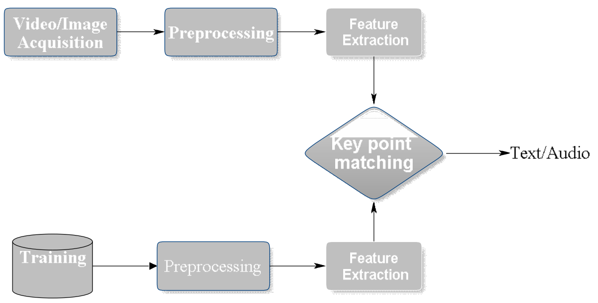


Fig 2. Block Diagram of Sign Language

-
Identification System

A general block diagram of sign language recognition system is shown in Fig.2. The recognition process is alienated into two phases- training and testing. In the training phase, the classifier has to be trained using the training dataset. The database can be either created by the researcher himself or an available database can be used. An external webcam, digital camera or inbuilt webcam in the laptops can be used to capture the training images. Most of the sign language recognition systems classify signs according to hand gestures only or in other words, facial expressions are omitted. The important steps involved in training phase are creation of database, preprocessing, feature extraction and training the classifier. The testing phase contains video/image acquisition (input can be videos or images), preprocessing, feature extraction and classification.

Image Preprocessing

In this Stage, the input gestures database is creating which contains the different gestures poses with many samples per gesture. Images of signs were resized to 256 by 256, by default MATLAB function "imresize" uses nearest neighbor interpolation to determine the values of pixels in the output image but other interpolation methods can be specified. Here "bicubic" method is used because if the specified output size is smaller than the size of the input image, "imresize" applies a low pass filter before interpolation to reduce aliasing.

Feature extraction using SIFT

The extracted features using this algorithm are invariant to image scaling, rotation, and partially invariant to illumination changes and affine or 3D projection [3]. These features share similar properties with neurons system that are used for object recognition in human vision. SIFT is divided into four stages respectively.

(1)Detection of Scale Space Local Extrema

The first stage of key point detection is to identify locations and scales that can be repeatable assigned under differing views of the Detecting locations that are invariant to scale change of the image can be accomplished by searching for stable features across all possible scales, using a continuous function of scale known as scale space[7]. Under a variety of Reasonable assumptions the only possible scale-space kernel is the Gaussian function. Therefore, the scale space of an image is defined as a function, $L(X,Y,\sigma)$ that is produced from the convolution of a variable-scale Gaussian, $G(X,Y,\sigma)$ with an input image, $I(X,Y)$:

-

$L(X,Y,\sigma) = G(X,Y, \sigma) * I(X,Y)$ .... (1)

To efficiently detect stable key point locations in same object. scale space, [8] proposed Using scale-space extrema in the difference-of-Gaussian function convolved with the image, $D(X,Y, \sigma)$ which can be computed from the difference of two nearby scales separated by a Constant multiplicative factor k:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$

$$= L(x, y, k\sigma) - L(x, y, \sigma) \quad ....(2)$$

(2) Locate maxima or minima in DOG image

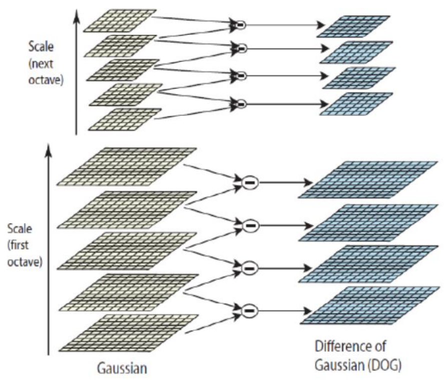


Fig.3 For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produced the set of scale space images shown on left.[3]

In this step we indelicately locate the maxima or minima. We go through each pixel and check all its neighbors. The check is done within the current image, and also the one above and below it. In Fig.4 X marks the current pixel and green circles mark the neighbors. This way, a total of 26 checks are made**.** X is marked as a "key point" if it is the greatest or least of all 26 neighbors. To increase chances of matching and stability of the algorithm, we also find sub pixel maxima or minima using Taylor series approximation.

-

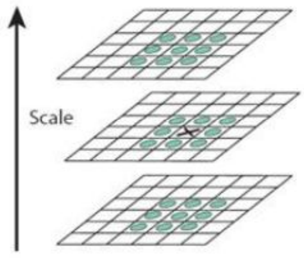


Fig.4 Maxima and minima of the difference-of-Gaussian images are detected by compairing a pixel (marked with X) to its 26 neighbours in 3x3 regions at the current and adjacent scales (marked with circles).

(3) Orientation assignment to the key point

Once the SIFT-feature location is determined, a

mainorientation is assigned to each feature based on local image gradients as shown in Fig.5. For each pixel of the region around the feature location the gradient magnitude and orientation arecomputed respectively as:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \quad .. (3)$$

$$\theta(x, y) = \tan^{-1}((L(x, y + 1) - L(x, y - 1))/(L(x + 1, y) - L(x - 1, y))) \quad ....(4)$$

The magnitude and orientation is calculated for all pixels around the key point. Then, histogram is created for this.

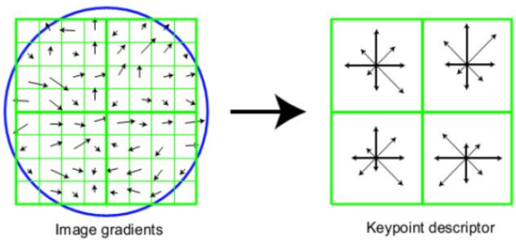


Fig.5. Orientation Assignment [3]

-

(4) Key point Descriptor

Its need to generate a very unique fingerprint for the key point also it to be relatively lenient when it is being compared against other key points. Fig.5 shows a 2x2 array of orientation histograms, whereas our experiment shows best results with a 4x4 array of histograms with 8 orientation bins in each. Therefore, the experiments in this paper take 4 x 4 x 8 = 128 element feature vector for each key point. These 128 numbers form the “feature vector”. The key point is uniquely identified by this feature vector.

EXPERIMENTAL RESULT

The hand gesture recognition system, presented in this paper is tested by using hand images from different people with varying shape, scale, rotation and size. The conclusions drawn based on the robustness of the features and Sign recognition. In our Experiment we take total 15 gestures C, R, S, 5 etc. of hand of different people. As shown Fig.6,7,8. When we calculate the feature descriptor of input image and it matches key points with the scaled image of character present in the database.

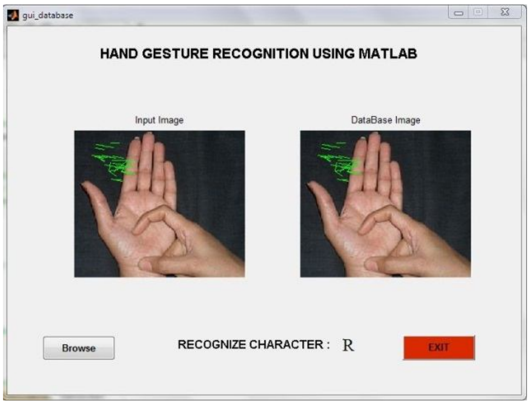


Fig.6 Matched Database image Vs. Input image of ISL character R

-

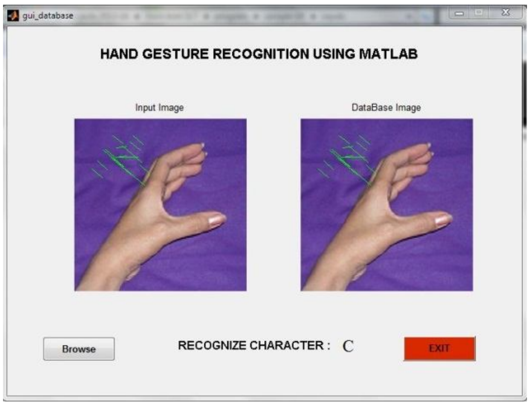


Fig.7 Matched Database image Vs. Input image of ASL character C.

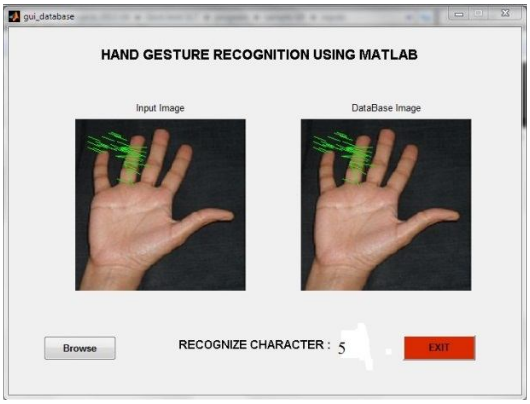


Fig.8 Matched Database image Vs. Input image of ASL character 5.

CONCLUSION

In this paper, we have deliberated the static signs of ISL and ASL from images or video sequences that have been recorded under controlled conditions. With the help of SIFT algorithm gestures were able to decode effectively. The features were extracted proficiently using SIFT. The Sift features were computed at the edges and they are invariant to image scaling, rotation, noise. They useful due to their distinctiveness. Also computation time was found to be less.

REFERENCES

1. Y. Wu, J. Lin, and T. Huang, "Capturing Natural Hand Articulation". In IEEE International Conference on Computer Vision II, 426–432, 2001.

-

2. H. Zhou and T. S. Huang, "Tracking articulated hand motion with Eigen dynamics analysis," in Proc. of International Conference on Computer Vision, vol. 2, 2003, pp. 1102–1109.

3. David G. Lowe, "Distinctive image features from scale-invariant key-points", International Journal of Computer Vision, Vol.60, No. 2, pp. 91-110, 2004

4. N. H. Dardas and N. D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques," *IEEE Transactions on Instrumentation and Measurement*, VOL.60. NO. 11, November 2011.

5. S. Pandita1, S. P. Narote "Hand Gesture Recognition using SIFT" *International Journal of Engineering Research & Technology (IJERT)* Vol. 2 Issue 1, January- 2013.

6. R. Gonzalez, R. Woods, and S. Eddins, Digital Image Processing Using MATLAB. Englewood Cliffs, NJ: Prentice-Hall, 2004.